

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 3 年   3 月 1 7 日  
Date of Application:

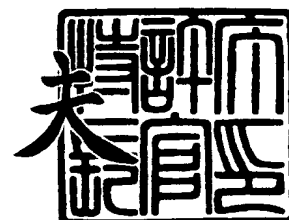
出 願 番 号                      特 願 2 0 0 3 - 0 7 2 3 2 4  
Application Number:  
[ST. 10/C]:                      [ J P 2 0 0 3 - 0 7 2 3 2 4 ]

出      願      人                      インターナショナル・ビジネス・マシーンズ・コーポレーシ  
Applicant(s):                      ョン

2 0 0 3 年   7 月 2 2 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 JP9030031

【提出日】 平成15年 3月17日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/28

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内

【氏名】 伊藤 晴美

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内

【氏名】 神山 淑朗

【特許出願人】

【識別番号】 390009531

【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

【代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【代理人】

【識別番号】 100108501

【弁理士】

【氏名又は名称】 上野 剛史

【復代理人】

【識別番号】 100104156

【弁理士】

【氏名又は名称】 龍華 明裕

【手数料の表示】

【予納台帳番号】 053394

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0207860

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 翻訳システム、辞書更新サーバ、翻訳方法、及び、これらのプログラムと記録媒体

【特許請求の範囲】

【請求項 1】 ドキュメントを翻訳する翻訳システムであって、  
予め定められた分野別に分類された複数の分野別辞書を管理する辞書管理部と、  
前記ドキュメントから、名詞句 (N o u n P h r a s e) を抽出するフレーズ抽出部と、  
前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、  
前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、  
前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部と  
を備える翻訳システム。

【請求項 2】 前記ドキュメントを翻訳したときに前記複数の分野別辞書のそれぞれを用いた頻度に基づいて、前記ドキュメントの分野を選択するドキュメント分野選択部を更に備え、  
前記登録先選択部は、前記ドキュメント分野選択部により選択された分野に基づいて、抽出された前記名詞句を登録すべき分野を選択する  
請求項 1 記載の翻訳システム。

【請求項 3】 前記ドキュメント分野選択部は、複数の前記ドキュメントのそれぞれの翻訳において、前記複数の分野別辞書のそれぞれを用いた頻度に基づいて、前記複数のドキュメントのそれぞれの分野を選択し、  
前記フレーズ抽出部は、前記複数のドキュメントから、前記名詞句を抽出し、  
前記登録先選択部は、前記複数のドキュメントのそれぞれについて、前記名詞句が出現した頻度及び当該ドキュメントの分野に基づいて、前記名詞句を登録すべき分野を選択する  
請求項 2 記載の翻訳システム。

【請求項 4】 前記複数のドキュメントにおいて前記名詞句が出現した頻度が予め定められた下限値未満の場合に、前記名詞句及び前記名詞句訳の組を、前記複数の分野別辞書のいずれにも登録させないことを選択する登録フレーズ選択部を更に備える請求項 3 記載の翻訳システム。

【請求項 5】 一の前記分野別辞書は、他の複数の前記分野別辞書に対応する複数の分野のいずれかに分類されない語句を登録する基本辞書であり、

前記登録先選択部は、

前記複数のドキュメントのそれぞれについて前記名詞句が出現した頻度及び当該ドキュメントの分野に基づいて、前記複数の分野のそれぞれに対する前記名詞句の出現頻度を算出する分野別出現頻度算出部と、

前記複数の分野のそれぞれに対する前記出現頻度に基づいて、前記名詞句及び前記名詞句訳の組を、前記複数の分野別辞書及び前記基本辞書のいずれに登録するかを選択する登録先分野選択部と

を有し、

前記辞書登録部は、前記登録先分野選択部により前記名詞句を前記基本辞書に登録することを選択された場合に、前記名詞句及び前記名詞句訳の組を前記基本辞書に登録する

請求項 2 記載の翻訳システム。

【請求項 6】 前記登録先選択部は、前記複数のドキュメントに対応する複数の分野のそれぞれに対して前記名詞句が出現した度合に基づいて、前記名詞句を登録すべき分野を選択する請求項 2 記載の翻訳システム。

【請求項 7】 前記翻訳部は、前記登録先選択部により選択された、前記名詞句を登録すべき分野に対応する前記分野別辞書を優先して前記名詞句を翻訳し、前記名詞句訳を生成する請求項 1 記載の翻訳システム。

【請求項 8】 前記翻訳部は、前記名詞句訳を生成する場合において、前記ドキュメントにおける名詞句以外の部分を翻訳する場合と比較し、名詞句として翻訳する文法規則を使用する優先度をより高く設定する請求項 1 記載の翻訳システム。

【請求項 9】 前記翻訳部は、

前記名詞句を翻訳して、前記名詞句訳の候補となる名詞句訳候補を生成する訳語生成部と、

前記名詞句訳候補を含むネットワーク上のページを検索するページ検索部と、  
前記名詞句訳候補を含むページがヒットしたか否かに基づいて、前記名詞句訳候補を前記名詞句訳とするか否かを選択する名詞句訳生成部と  
を有する請求項 1 記載の翻訳システム。

【請求項 10】 予め定められた分野別に分類された複数の分野別辞書を管理する端末においてドキュメントの翻訳に用いる辞書を更新する辞書更新サーバであって、

前記ドキュメントから抽出された名詞句を前記端末から受信する名詞句受信部と、

前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、

前記端末から受信した前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、

前記前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部と

を備える辞書更新サーバ。

【請求項 11】 ドキュメントを翻訳し、翻訳に用いる辞書を外部の辞書更新サーバの指示に基づき更新する端末を有する翻訳システムであって、

前記端末は、

予め定められた分野別に分類された複数の分野別辞書を格納する翻訳辞書記録部と、

前記ドキュメントから名詞句を抽出するフレーズ抽出部と、

前記複数の分野別辞書を用いて前記ドキュメントを翻訳するドキュメント翻訳部と

を備え、

前記辞書更新サーバは、

前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記

名詞句を登録すべき分野を選択する登録先選択部と、

前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、

前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句および前記名詞句訳の組の登録を指示する辞書登録部と

を備え、

前記翻訳辞書記録部は、前記辞書登録部の登録指示に基づき、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する翻訳システム。

【請求項 12】 ドキュメントを翻訳する翻訳システム用のプログラムであって、

前記翻訳システムを、

予め定められた分野別に分類された複数の分野別辞書を管理する辞書管理部と

前記ドキュメントから、名詞句 (N o u n P h r a s e) を抽出するフレーズ抽出部と、

前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、

前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、

前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部と

して機能させるプログラム。

【請求項 13】 予め定められた分野別に分類された複数の分野別辞書を管理する端末においてドキュメントの翻訳に用いる辞書を更新する辞書更新サーバ用のプログラムであって、

当該プログラムは、当該辞書更新サーバを、

前記ドキュメントから抽出された名詞句を前記端末から受信する名詞句受信部と、

前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、

前記端末から受信した前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、

前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部と

して機能させるプログラム。

【請求項 14】 請求項 12 又は請求項 13 のいずれかに記載のプログラムを記録した記録媒体。

【請求項 15】 コンピュータによりドキュメントを翻訳する翻訳システムにおける翻訳方法であって、

予め定められた分野別に分類された複数の分野別辞書を、コンピュータにより管理させる辞書管理段階と、

前記ドキュメントから、名詞句 (N o u n P h r a s e) を、コンピュータにより抽出させるフレーズ抽出段階と、

前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を、コンピュータにより選択させる登録先選択段階と、

前記名詞句を翻訳して前記名詞句の訳である名詞句訳を、コンピュータにより生成させる翻訳段階と、

前記登録先選択段階により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を、コンピュータにより登録させる辞書登録段階とを備える翻訳方法。

#### 【発明の詳細な説明】

##### 【0001】

#### 【発明の属する技術分野】

本発明は、翻訳システム、辞書更新サーバ、翻訳方法、及び、これらのプログラムと記録媒体に関する。特に本発明は、ドキュメントの翻訳に用いる翻訳辞書を自動的に更新する翻訳システム、辞書更新サーバ、翻訳方法、及び、これらのプログラムと記録媒体に関する。

##### 【0002】

#### 【従来の技術】



従来、ドキュメントを翻訳する翻訳システムにおいて翻訳の精度を高めるために利用できる技術として、以下の3つの技術が開示されている。

第1の技術は、操作対象として指定された語を基に辞書の見出し語となる見出し文字列を生成し、辞書に登録する方法を開示する（特許文献1参照。）。

第2の技術は、第1の言語のテキストと、第1の言語のテキストを第2の言語に翻訳した対訳テキストとから各語の対訳データを抽出し、対訳辞書を生成する方法を開示する（特許文献2参照。）。

第3の技術は、ある訳例と、この訳例の一部の語が変化した他の訳例とに基づいて、訳文の一部が変数化された訳例辞書を生成する方法を開示する（特許文献3参照。）。

【0003】

【特許文献1】

特開平6-28391号公報

【0004】

【特許文献2】

特開平9-128396号公報

【0005】

【特許文献3】

特開2002-297588号公報

【0006】

【発明が解決しようとする課題】

近年の技術発展や、ビジネスのグローバル化に伴い、日々多くの語句が作られ、その一部は急速に普及している。このような状況の下、ドキュメントを翻訳する翻訳システムにおいては、新しい語句に対応する訳をユーザ等が登録しなければ、翻訳対象の語句が翻訳辞書に登録されている可能性が低下する結果、翻訳精度が低下してしまうという問題がある。

【0007】

上記の特許文献1から3に示される技術は、新しい語句に対応する訳を辞書に登録する機能を提供しないため、以上の問題を解決することができなかった。

## 【0008】

そこで本発明は、上記の課題を解決することのできる翻訳システム、辞書更新サーバ、翻訳方法、及び、これらのプログラムと記録媒体を提供することを目的とする。この目的は特許請求の範囲における独立項に記載の特徴の組み合わせにより達成される。また従属項は本発明の更なる有利な具体例を規定する。

## 【0009】

## 【課題を解決するための手段】

即ち、本発明の第1の形態によると、ドキュメントを翻訳する翻訳システムであって、予め定められた分野別に分類された複数の分野別辞書を管理する辞書管理部と、前記ドキュメントから、名詞句 (Noun Phrase) を抽出するフレーズ抽出部と、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部とを備える翻訳システム、当該翻訳システムを構成する辞書更新サーバと端末、翻訳方法、及びこれらのプログラム、記録媒体を提供する。

## 【0010】

なお上記の発明の概要は、本発明の必要な特徴の全てを列挙したものではなく、これらの特徴群のサブコンビネーションも又発明となりうる。

## 【0011】

## 【発明の実施の形態】

以下、発明の実施の形態を通じて本発明を説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、又実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

## 【0012】

図1は、本実施形態に係る翻訳システム10の構成を示す。本実施形態に係る翻訳システム10は、翻訳処理中に未知のフレーズ (Phrase、句) を抽出し、抽出したフレーズの訳を翻訳により生成する。そして、フレーズ及びフレー

ズの訳であるフレーズ訳を組として翻訳辞書に登録することにより自律的に語彙拡充を行なって、未知の語句の増加に伴う翻訳精度の低下を抑えることを目的とする。

#### 【0013】

翻訳システム10は、ドキュメントを翻訳し、翻訳した結果である翻訳ドキュメントを出力する翻訳フロントエンドシステム100と、翻訳フロントエンドシステム100により抽出されたフレーズの訳を生成して翻訳フロントエンドシステム100の翻訳辞書を更新する辞書更新サーバ160とを備える。

#### 【0014】

翻訳フロントエンドシステム100は、利用者又は他のアプリケーションプログラム等により指定されたドキュメントを翻訳する。翻訳フロントエンドシステム100は、翻訳結果を利用する利用者のパーソナルコンピュータ、PDA、又は携帯電話等の端末上で実現されてもよい。これに代えて、翻訳フロントエンドシステム100は、利用者がブラウザ等を用いて通信ネットワークを介してアクセスするサーバ上で実現されてもよく、この場合において辞書更新サーバ160と同一のサーバ上に実現されてもよい。

#### 【0015】

翻訳フロントエンドシステム100は、翻訳辞書記録部110と、ドキュメント翻訳部120と、抽出フレーズ記録部125とを有する。

翻訳辞書記録部110は、翻訳フロントエンドシステム100による翻訳に用いられる翻訳辞書117を格納する。翻訳辞書117は、複数の分野にそれぞれ対応する複数の分野別辞書115a～bを含む。分野別辞書115bは、予め定められた分野、すなわち例えばスポーツ、家庭、ビジネス、及び科学等の分野別に分類される語句等の訳を登録する。分野別辞書115aは、他の分野別辞書115、すなわち複数の分野別辞書115bに対応する複数の分野のいずれかに分類されない語句等とその訳を登録する。すなわち、分野別辞書115aは、複数の分野別辞書115bに対応する複数の分野のいずれにも属さない分野である基本分野に属する語句等を登録する。分野別辞書115は、翻訳対象のドキュメントの分野に応じて、少なくとも1つが優先的に用いられてよい。また、分野別辞

書 115 a～b は、翻訳フロントエンドシステム 100 による翻訳に用いられる文法規則を格納する文法辞書として更に機能してもよい。

#### 【0016】

ドキュメント翻訳部 120 は、英語等の第 1 の言語で記述されたドキュメントを、日本語等の第 2 の言語で記述された翻訳ドキュメントに翻訳する。また、翻訳処理においてドキュメントから未知のフレーズを抽出し、ドキュメントの分野と共に抽出フレーズ記録部 125 へ出力する。ここで、ドキュメント翻訳部 120 は、ドキュメントの内容等に基づいて、ドキュメントの分野を選択してもよく、これに代えて、利用者からの指定に基づいてドキュメントの分野を設定してもよい。

#### 【0017】

抽出フレーズ記録部 125 は、ドキュメント翻訳部 120 によりドキュメントから抽出されたフレーズを、フレーズの出現分野に対応付けて格納する。ここで、フレーズの出現分野とは、当該フレーズが出現したドキュメントの分野である。そして、抽出フレーズ記録部 125 は、格納したフレーズ及び出現分野の組を、例えば定期的又は予め定められたタイミングで、辞書更新サーバ 160 へ送信する。

#### 【0018】

辞書更新サーバ 160 は、翻訳フロントエンドシステム 100 の抽出フレーズ記録部 125 から受信したフレーズの訳を生成して、翻訳フロントエンドシステム 100 の翻訳辞書を更新する。辞書更新サーバ 160 は、翻訳結果を利用する利用者の端末上に、翻訳フロントエンドシステム 100 と共に実現されてもよい。これに代えて、辞書更新サーバ 160 は、利用者がブラウザ等を用いて通信ネットワークを介してアクセスするサーバ上に、翻訳フロントエンドシステム 100 と共に実現されてもよく、通信ネットワークを介して翻訳フロントエンドシステム 100 が実装されたサーバと通信する他のサーバ上に実現されてもよい。

#### 【0019】

辞書更新サーバ 160 は、フレーズ受信部 127 と、フレーズ分類部 130 と、登録フレーズ記録部 140 と、翻訳辞書記録部 170 と、フレーズ翻訳部 18

0と、更新辞書185と、辞書登録部190とを有する。

#### 【0020】

フレーズ受信部127は、翻訳対象のドキュメントから抽出されたフレーズを抽出フレーズ記録部125から受信する。フレーズ分類部130は、抽出フレーズ記録部125からフレーズ受信部127を介して受信したフレーズのうち、翻訳辞書117に登録すべきフレーズを選択し、当該フレーズに登録すべき登録分野を選択する。そして、フレーズを分野別辞書115aに登録する場合には、当該フレーズ及び登録分野（基本分野）を登録フレーズ記録部140上の分野別登録フレーズ記録ファイル145aとして格納し、フレーズを分野別辞書115bに登録する場合には、当該フレーズ及び登録分野を登録フレーズ記録部140上の分野別登録フレーズ記録ファイル145bとして格納する。登録フレーズ記録部140は、分野別登録フレーズ記録ファイル145a～bに格納された、登録すべきフレーズ及び当該フレーズの登録分野を、フレーズ翻訳部180へ供給する。

#### 【0021】

翻訳辞書記録部170は、翻訳辞書記録部110と同一の機能を有し、翻訳フロントエンドシステム100から受信したフレーズの翻訳に用いられる翻訳辞書177を格納する。ここで、翻訳辞書177に含まれる分野別辞書175a～bは、分野別辞書115a～bと同期して更新されることにより、同一の内容に保たれてもよい。これに代えて、翻訳フロントエンドシステム100又は辞書更新サーバ160は、分野別辞書175a～bの一部の内容を、分野別辞書115a～bに登録してもよい。また、翻訳フロントエンドシステム100及び辞書更新サーバ160が同一の端末又はサーバ上に設けられる場合等において、翻訳辞書記録部170に代えて、翻訳辞書記録部110が直接フレーズ翻訳部180に接続され、フレーズ翻訳部180が翻訳辞書記録部110を直接用いる構成を採ってもよい。

#### 【0022】

フレーズ翻訳部180は、本発明に係る翻訳部の一例であり、翻訳フロントエンドシステム100から受信した各フレーズを翻訳してフレーズ訳を生成し、翻

訳辞書 117 及び翻訳辞書 177 の更新に用いる更新辞書を生成する。更新辞書 185 は、フレーズ翻訳部 180 が生成した更新辞書を格納する。辞書登録部 190 は、更新辞書 185 の更新辞書に基づいて、翻訳辞書 117 及び翻訳辞書 177 にフレーズ及びフレーズ訳の組を登録する。ここで、辞書登録部 190 は、当該フレーズの登録分野に対応する分野別辞書 115 b 及び分野別辞書 175 b に、当該フレーズ及びフレーズ訳の組を登録する。一方、当該フレーズの登録分野が複数の分野別辞書 115 b のいずれにも分類されない場合、すなわち当該フレーズの登録分野が基本分野である場合には、基本分野の分野別辞書 115 a 及び分野別辞書 175 a に当該フレーズ及びフレーズ訳の組を登録する。

#### 【0023】

なお、辞書登録部 190 は、いずれかの分野別辞書 115 に当該フレーズ及びフレーズ訳の組を登録する場合、翻訳辞書記録部 110 に対してフレーズ及びフレーズ訳の組の登録を指示する。そして、翻訳辞書記録部 110 は、辞書登録部 190 の登録指示に基づき、選択された分野に対応する分野別辞書 115 に、フレーズ及びフレーズ訳の組を登録する。

#### 【0024】

以上に示した翻訳システム 10 によれば、翻訳フロントエンドシステム 100 の翻訳対象となったドキュメントから抽出したフレーズの訳を辞書更新サーバ 160 が生成し、当該フレーズに対応する分野の分野別辞書 115 に登録することができる。これにより、翻訳対象となったドキュメントの分野に対応する翻訳辞書の語彙拡充を行なうことができ、未知の語句の増加に伴う翻訳精度の低下を抑えることができる。

#### 【0025】

なお、以上において、翻訳フロントエンドシステム 100 及び辞書更新サーバ 160 は、図 1 と異なる構成部材の組合せにより実現されてもよい。例えば、フレーズ分類部 130 及び登録フレーズ記録部 140 は、辞書更新サーバ 160 に代えて翻訳フロントエンドシステム 100 の構成部材として実現されてもよい。

#### 【0026】

図 2 は、本実施形態に係る翻訳辞書記録部 110 及び翻訳辞書記録部 170 に

格納される翻訳辞書 117 及び翻訳辞書 177 の階層構造の一例を示す。階層構造の最上位には、翻訳辞書 117 及び翻訳辞書 177 に対応する辞書 900 が位置する。辞書 900 は、分野別に分類された分野別辞書 910 と、分野別辞書 910 に対応する複数の分野のいずれかに分類されない語句を登録する基本辞書 905 とに分類される。分野別辞書 910 は、例えばスポーツや家庭等の複数の主分野 915 に分類される。そして、各主分野 915 に対応して、当該主分野 915 を更に細分化した分野である副分野に対応する副分野別辞書 925 と、当該主分野 915 においていずれの副分野にも属さない語句又はフレーズ等を登録する主分野別辞書 920 とが設けられる。

#### 【0027】

分野別辞書 115b 及び分野別辞書 175b のそれぞれは、副分野別辞書 925 に対応してよい。この場合において、複数の分野別辞書 115b 及び複数の分野別辞書 175b に対応する複数の分野のいずれかに分類されない語句を登録する分野別辞書 115a 及び分野別辞書 175a は、主分野別辞書 920 に対応してもよく、これに代えて基本辞書 905 に対応してもよい。

#### 【0028】

また、分野別辞書 115b 及び分野別辞書 175b のそれぞれは、いずれかの主分野 915 に対応する複数の副分野別辞書 925 及び主分野別辞書 920 を含んでもよい。この場合において、分野別辞書 115a 及び分野別辞書 175a は、基本辞書 905 に対応してもよい。

#### 【0029】

図 3 は、本実施形態に係るドキュメント翻訳部 120 の構成を示す。ドキュメント翻訳部 120 は、辞書管理部 200 と、形態素解析部 210 と、フレーズ抽出部 220 と、構文解析部 230 と、ドキュメント訳生成部 240 と、ドキュメント分野選択部 250 とを含む。

#### 【0030】

辞書管理部 200 は、翻訳辞書記録部 110 に記録された複数の分野別辞書 115 を管理する。形態素解析部 210 は、ドキュメントに含まれる各文について、形態素解析を行なう。フレーズ抽出部 220 は、形態素解析の結果に基づいて

、フレーズをドキュメントから抽出する。構文解析部 230 は、形態素解析の結果に基づいて、ドキュメントに含まれる各文の構文解析を行う。ドキュメント訳生成部 240 は、形態素解析結果、及び構文解析結果に基づいて、辞書管理部 200 を介して複数の分野別辞書 115 を参照し、ドキュメントを翻訳した翻訳ドキュメントを生成する。ドキュメント分野選択部 250 は、ドキュメント訳生成部 240 がドキュメントを翻訳したときに複数の分野別辞書 115 のそれぞれを用いた頻度に基づいて、ドキュメントの分野を選択する。

### 【0031】

図 4 は、本実施形態に係るドキュメント翻訳部 120 の処理フローを示す。

まず、形態素解析部 210 は、ドキュメントに含まれる各文について、当該文章を構成する、意味を持つ最小単位である形態素 (morpheme) を解析し、単語を認識する (S300)。この処理において、形態素解析部 210 は、分野別辞書 115 a～b に格納された文法規則を参照し、文法規則に基づく形態素解析を行なう。

### 【0032】

次に、フレーズ抽出部 220 は、形態素解析の結果に基づいて、未知のフレーズをドキュメントから抽出する (S320)。本実施形態においてフレーズ抽出部 220 は、翻訳辞書 117 に未登録の名詞句 (Noun Phrase) をドキュメントから抽出する。これに代えて、フレーズ抽出部 220 は、動詞句等を含む各種のフレーズを抽出してもよい。なお、S320 において、フレーズ抽出部 220 は、形態素解析の結果に基づいて認識されるフレーズの訳が複数の分野別辞書 115 に登録されていない場合に、当該フレーズが未知であると判断する。

### 【0033】

次に、構文解析部 230 は、形態素解析の結果に基づいて、ドキュメントに含まれる各文の構文を解析する (S330)。次に、ドキュメント訳生成部 240 は、形態素解析の結果、及び構文解析の結果に基づき認識されるドキュメント内の単語及びフレーズ等の複合語のそれぞれについて、辞書管理部 200 を介して複数の分野別辞書 115 を参照して、単語の訳やフレーズ訳等の訳語を生成する



翻訳処理を行う (S 3 4 0)。

#### 【0034】

次に、ドキュメント分野選択部 2 5 0 は、ドキュメント訳生成部 2 4 0 がドキュメントを翻訳したときに複数の分野別辞書 1 1 5 のそれぞれを用いた頻度に基づいて、ドキュメントの分野を選択する (S 3 5 0)。すなわち例えば、ドキュメント分野選択部 2 5 0 は、複数の分野別辞書 1 1 5 のそれぞれを用いた回数を、複数の分野別辞書 1 1 5 のいずれかを用いた回数で割った値を、当該分野の単語又はフレーズ等がドキュメントに含まれる頻度として算出する。そして、ドキュメント分野選択部 2 5 0 は、ある分野の単語又はフレーズ等が、他の分野の単語又はフレーズ等と比較しドキュメント中に頻出する場合に、この分野を当該ドキュメントの分野として選択する。この処理の一例として、ドキュメント分野選択部 2 5 0 は、上記の頻度が予め定められたしきい値以上となる分野が存在する場合に、この分野を当該ドキュメントとして選択してもよい。

#### 【0035】

次に、ドキュメント分野選択部 2 5 0 は、当該ドキュメントの分野を、ドキュメントから抽出された複数のフレーズの出現分野とし、ドキュメントから抽出されたフレーズ及び当該フレーズの出現分野の組を、抽出フレーズ記録部 1 2 5 へ登録する (S 3 5 5)。ここで、ドキュメント分野選択部 2 5 0 は、翻訳対象の 1 又は複数のドキュメントにおいてそれぞれのフレーズが出現した回数を、当該フレーズが出現した頻度として、当該フレーズに対応付けて抽出フレーズ記録部 1 2 5 へ登録する。そして、ドキュメント訳生成部 2 4 0 は、ドキュメントの分野に対応する分野別辞書 1 1 5 を優先して、ドキュメントを翻訳する (S 3 6 0)。

#### 【0036】

翻訳対象のドキュメントが複数存在する場合、すなわち例えば利用者が複数のドキュメントを順次翻訳させる場合等において、ドキュメント翻訳部 1 2 0 は、S 3 0 0 から S 3 6 0 に示した処理を複数のドキュメントのそれぞれについて実行する (S 3 7 0)。この結果、複数のドキュメントのそれぞれの翻訳において、形態素解析部 2 1 0 は、複数のドキュメントのそれぞれについて形態素解析を

行い、フレーズ抽出部 220 は、複数のドキュメントのそれぞれからフレーズを 1 以上抽出し、構文解析部 230 は、複数のドキュメントのそれぞれについて構文解析を行う。そして、ドキュメント訳生成部 240 は、複数のドキュメント内の単語及び複合語のそれぞれについて訳語を生成し、ドキュメント分野選択部 250 は、複数の分野別辞書 115 のそれぞれを用いた頻度に基づいて、複数のドキュメントのそれぞれの分野を選択する。

#### 【0037】

以上に示したドキュメント翻訳部 120 によれば、辞書更新サーバ 160 が新たなフレーズ及びフレーズ訳を複数の分野別辞書 115 に順次登録することにより、構文解析部 230 は、当該フレーズを構成する各語の構文を解析することなく当該フレーズを句として認識することができる。この結果、ドキュメント翻訳部 120 における構文解析の精度を高めると共に、構文解析の速度を上げることができる。

#### 【0038】

図 5 は、本実施形態に係るフレーズ分類部 130 の構成を示す。フレーズ分類部 130 は、登録フレーズ選択部 400 と、登録先選択部 410 とを有する。

#### 【0039】

登録フレーズ選択部 400 は、翻訳対象の 1 又は複数のドキュメント中に各フレーズが出現した頻度に基づき、各フレーズを翻訳辞書に登録すべきか否かを選択する。登録先選択部 410 は、フレーズ抽出部 220 により抽出され、登録フレーズ選択部 400 により登録すべきことが選択されたフレーズのそれぞれについて、複数の分野別辞書 115 にそれぞれ対応する複数の分野のうち、当該フレーズに登録すべき分野を選択する。登録先選択部 410 は、分野別出現頻度算出部 420 と、登録先分野選択部 430 とを含む。

#### 【0040】

分野別出現頻度算出部 420 は、翻訳対象の 1 又は複数のドキュメントのそれぞれについて、各フレーズが出現した頻度及び当該ドキュメントの分野に基づいて、複数の分野に対する当該フレーズの出現頻度を算出する。登録先分野選択部 430 は、各フレーズについて、複数の分野に対する当該フレーズの出現頻度に

基づいて、当該フレーズを複数の分野別辞書 115 のいずれに登録するかを選択する。

#### 【0041】

図 6 は、本実施形態に係るフレーズ分類部 130 の処理フローを示す。

まず、登録フレーズ選択部 400 は、抽出フレーズ記録部 125 から受信した 1 以上のフレーズを、分野毎の出現頻度順に並べ替える (S500)。次に、登録フレーズ選択部 400 は、各フレーズについて、翻訳対象の 1 又は複数のドキュメントにおいて当該フレーズが出現した頻度が予め定められた下限値未満の場合に、当該フレーズ及び当該フレーズのフレーズ訳の組を、複数の分野別辞書 115 のいずれにも登録させないことを選択する (S505)。より具体的には、登録フレーズ選択部 400 は、抽出フレーズ記録部 125 から受信した 1 以上のフレーズのうち、いずれの分野別辞書 115 にも登録させないことを選択したフレーズに関する情報を除いた情報を登録先選択部 410 に供給する。

#### 【0042】

次に、分野別出現頻度算出部 420 は、翻訳対象の 1 又は複数のドキュメントのそれぞれについてフレーズが出現した頻度及び当該ドキュメントの分野に基づいて、複数の分野のそれぞれに対する当該フレーズの出現頻度を算出する (S515)。

#### 【0043】

次に、登録先分野選択部 430 は、複数の分野のそれぞれに対する当該フレーズの出現頻度に基づいて、当該フレーズ及び当該フレーズのフレーズ訳の組を分野別辞書 115 a 及び複数の分野別辞書 115 b のいずれに登録するかを選択する。より具体的には、当該フレーズがある特定の分野に頻出する場合 (S520)、フレーズを登録すべき分野としてこの特定分野を選択し、当該フレーズをこの特定分野に対応付けて、分野別登録フレーズ記録ファイル 145 b に格納する (S530)。当該フレーズがいずれかの分野に偏って頻出しない場合 (S520)、フレーズおよびフレーズの訳の組を基本辞書である分野別辞書 115 a に登録することを選択し、当該フレーズを基本分野に対応付けて、分野別登録フレーズ記録ファイル 145 a に格納する (S535)。フレーズ分類部 130 は、

以上の S 5 0 5 から S 5 3 5 に示した処理を、抽出フレーズ記録部 1 2 5 から受信した全てのフレーズに対して行う (S 5 4 0)。

#### 【0044】

以上に示した処理により、フレーズ分類部 1 3 0 は、1 又は複数のドキュメントのそれぞれについて、当該ドキュメント中にフレーズが出現した頻度と、ドキュメント分野選択部 2 5 0 により選択された当該ドキュメントの分野とに基づいて、1 又は複数のドキュメントから抽出されたフレーズを登録すべき分野を選択する。

#### 【0045】

例えば、分野 C 1 のドキュメント D 1 と、分野 C 2 のドキュメント D 2 及び D 3 について、フレーズ A がそれぞれ d 1、d 2、及び d 3 の出現頻度で出現した場合、フレーズ分類部 1 3 0 は、次に例示する方法でフレーズを登録すべき分野を選択してよい。登録フレーズ選択部 4 0 0 は、フレーズ A の出現頻度 (d 1 + d 2 + d 3) がフレーズ A を登録する条件を満たさない場合、当該フレーズをいずれの分野別辞書 1 1 5 にも登録させないことを選択する。フレーズ A を分野別辞書 1 1 5 に登録する場合において、分野別出現頻度算出部 4 2 0 は、フレーズ A の分野 C 1 における出現頻度 d 1、及び分野 C 2 における出現頻度 (d 2 + d 3) を算出する。登録先分野選択部 4 3 0 は、出現頻度 d 1 と出現頻度 (d 2 + d 3) に基づいて、フレーズ A がいずれの分野に偏って頻出するかを判断し、フレーズ A をいずれの分野別辞書 1 1 5 に登録するかを選択する。

#### 【0046】

以上に示したフレーズ分類部 1 3 0 によれば、あるフレーズが 1 又は複数のドキュメント中に下限値以上の頻度で出現した場合に、当該フレーズを翻訳辞書 1 1 7 に登録することを選択する。これにより、翻訳システム 1 0 は、出現頻度が低く、分野の分類を十分正確に行なえないフレーズを翻訳辞書 1 1 7 に登録しないため、翻訳の精度が低下するのを防ぐことができる。また、フレーズ分類部 1 3 0 によれば、ある特定分野で頻出するフレーズを、この特定分野に対応する分野別辞書 1 1 5 b に登録することを選択する一方、いずれかの分野に偏って出現しないフレーズを、いずれの分野にも分類されない分野別辞書 1 1 5 a に登録す

ることを選択する。これにより、翻訳システム 10 は、フレーズが出現する分野に応じて当該フレーズを適切な分野別辞書 115 に登録することができ、翻訳辞書 117 の語彙拡充を適切に行なう結果翻訳の精度を向上することができる。

#### 【0047】

次に、より具体的なフレーズ分類部 130 の処理の一例を示す。

まず、登録フレーズ選択部 400 は、抽出フレーズ記録部 125 から受信したフレーズ及びフレーズの出現分野に基づき、フレーズ  $t_i$  が出現分野  $d_j$  に出現した頻度（回数）を示す、以下の行列を生成する（式（1））。

#### 【0048】

【数 1】

$$A = \begin{matrix} & d_1 & d_2 & d_3 & d_4 \\ \begin{matrix} t_1 \\ t_2 \\ t_3 \end{matrix} & \begin{bmatrix} 10 & 0 & 0 & 1 \\ 1 & 12 & 0 & 1 \\ 3 & 5 & 3 & 2 \end{bmatrix} \end{matrix} \quad (1)$$

#### 【0049】

ここで、各フレーズ  $t_i$  を、それぞれの出現分野に出現した頻度を要素とするベクトル  $t_i$  とすると、フレーズ  $t_i$  が出現分野  $d_j$  に出現した度合は、一例として以下の式（2）に示す、フレーズ  $t_i$  の出現分野  $d_j$  に対する類似度により表すことができる。

#### 【0050】

【数 2】

$$\text{sim}(\vec{t_i}, \vec{e_j}) = \frac{\vec{t_i} \cdot \vec{e_j}}{|\vec{t_i}| \cdot |\vec{e_j}|} \quad (2)$$

#### 【0051】

分野別出現頻度算出部 420 は、フレーズ  $t_i$  の出現分野  $d_j$  に対する出現頻

度として、例えば以下の式 (3) の  $tf(i, j)$  に示す様に、最大の出現頻度を用いて正規化した出現頻度を算出する。

【0052】

【数3】

$$tf_{(i,j)} = K + (1-K) \frac{A_{(i,j)}}{\max_{i,j} (A_{(i,j)})} \quad (3)$$

【0053】

なお、式 (3) において、 $K$  は、出現頻度が当該フレーズの登録有無の判断に与える影響を定める定数である。

【0054】

登録先分野選択部 430 は、フレーズ  $t_i$  が出現分野  $d_j$  に出現した度合、及び／又は、フレーズ  $t_i$  の出現分野  $d_j$  に対する出現頻度に基づいて、フレーズ  $t_i$  を出現分野  $d_j$  に登録するか否かを選択する。ここで、登録先分野選択部 430 は、フレーズ  $t_i$  が出現分野  $d_j$  に出現した度合、及び、フレーズ  $t_i$  の出現分野  $d_j$  に対する出現頻度に基づいて、フレーズ  $t_i$  を出現分野  $d_j$  に登録するか否かを選択する場合、式 (2) の類似度と、式 (3) の出現頻度の積に基づいて、フレーズ  $t_i$  を出現分野  $d_j$  に登録するか否かを定めてよい。

【0055】

フレーズ分類部 130 は、まず、複数の副分野について上記の処理を行うことにより、特定の副分野に偏って出現し、かつ、出現頻度の高いフレーズから順に、翻訳辞書 117 及び翻訳辞書 177 に登録することができる。そして、この処理によっていずれかの副分野に登録されるフレーズを除き、複数の主分野について再度上記の処理を行うことにより、特定の副分野に偏らず、特定の主分野に偏って出現し、かつ、出現頻度の高いフレーズから順に翻訳辞書 117 及び翻訳辞書 177 に登録することができる。

【0056】

なお、他の実施形態として、登録先分野選択部 430 は、上記の S520、S530、及び S535 において、ある特定の分野におけるフレーズの出現頻度が予め定められた値以上である場合に、フレーズを登録すべき分野としてこの特定分野を選択し、ある特定分野におけるフレーズの出現頻度が予め定められた値未満である場合に、フレーズおよびフレーズの訳の組を基本辞書である分野別辞書 115a に登録することを選択してもよい。

#### 【0057】

図 7 は、本実施形態に係るフレーズ翻訳部 180 の構成を示す。フレーズ翻訳部 180 は、優先度設定部 605 と、訳語生成部 600 と、ページ検索部 610 と、形態素解析部 613 と、構文解析部 616 と、フレーズ訳生成部 620 とを含む。

#### 【0058】

優先度設定部 605 は、分野別登録フレーズ記録ファイル 145.a～b に格納された各フレーズのそれぞれについて、当該フレーズの翻訳において優先して使用する分野別辞書 175 を選択する。訳語生成部 600 は、分野別登録フレーズ記録ファイル 145.a～b に格納された各フレーズのそれぞれについて、当該フレーズを翻訳して、フレーズ訳の候補となるフレーズ訳候補を生成する。ページ検索部 610 は、各フレーズのそれぞれに対応するフレーズ訳候補を含むネットワーク上のページを検索する。形態素解析部 613 は、形態素解析部 210 と同様の構成及び機能を有し、対象のフレーズの形態素解析を行なう。構文解析部 616 は、構文解析部 230 と同様の構成及び機能を有し、対象のフレーズの構文解析を行なう。フレーズ訳生成部 620 は、形態素解析及び構文解析の結果又はページ検索部 610 によるページの検索結果に基づいて、当該フレーズのフレーズ訳を生成する。

#### 【0059】

図 8 は、本実施形態に係るフレーズ翻訳部 180 の処理フローを示す。まず、優先度設定部 605 は、登録フレーズ記録部 140 内の分野別登録フレーズ記録ファイル 145.a～b に格納された、翻訳辞書 117 に登録すべきフレーズを順次取得する。取得したフレーズがいずれかの分野別登録フレーズ記録ファイル 1

45bに格納されたフレーズである場合(S700)、登録先分野選択部430により選択され、当該フレーズに対応付けて分野別登録フレーズ記録ファイル145に格納された、当該フレーズを登録すべき登録分野に対応する分野別辞書175bの優先度を、他の分野別辞書の優先度と比較し高める(S710)。これにより、優先度設定部605は、当該分野別辞書175bを優先して使用することを決定する(S710)。一方、取得したフレーズが分野別登録フレーズ記録ファイル145aに格納されたフレーズである場合(S700)、全ての分野別辞書175bを同一の優先度で使用することを決定する。

#### 【0060】

次に、形態素解析部613、構文解析部616及びフレーズ訳生成部620は、対象のフレーズを翻訳してフレーズの訳であるフレーズ訳を生成する(S720)。すなわち、形態素解析部613は、分野別辞書175a～bを参照して、対象のフレーズの形態素解析を行なう。次に、構文解析部616は、形態素解析の結果に基づいて、対象のフレーズの構文解析を行なう。フレーズ訳生成部620は、形態素解析及び構文解析の結果に基づき認識されるドキュメント内の単語等のそれぞれについて、分野別辞書175a～bを参照して、対象のフレーズを翻訳し、フレーズ訳を生成する。ここで、S710において一の分野別辞書175bを優先して使用することを決定した場合、フレーズ訳生成部620は、当該分野別辞書175bを優先して当該フレーズを翻訳し、フレーズ訳を生成する。

#### 【0061】

次に、フレーズ訳生成部620は、対象のフレーズと、当該フレーズのフレーズ訳とに基づき、翻訳辞書117の更新に用いる更新辞書を生成する(S730)。ここで、フレーズ訳生成部620が生成する更新辞書は、当該フレーズに対応付けて、当該フレーズのフレーズ訳及び当該フレーズを登録すべき登録分野又は当該フレーズを登録すべき分野別辞書115を識別する識別情報を保持する。

#### 【0062】

フレーズ翻訳部180は、分野別登録フレーズ記録ファイル145a～bに格納された、翻訳辞書117に登録すべき各フレーズについて、S700からS730の処理を行う(S740)。



**【0063】**

以上に示したフレーズ翻訳部180によれば、訳語生成部600及びフレーズ訳生成部620は、登録すべきフレーズのフレーズ訳を生成するときに、当該フレーズの登録分野に対応する分野別辞書175を優先して用いる。これにより、当該フレーズを登録すべき分野で用いられる語句の訳を優先して使用したフレーズ訳を生成することができ、フレーズの翻訳精度を高めることができる。

**【0064】**

図9は、本実施形態に係るフレーズ翻訳部180の、ネットワークを介したフレーズ訳生成処理フローを示す。フレーズ翻訳部180は、インターネット等のネットワーク上のページを利用してフレーズ訳を生成する場合、図8のS720において図9に示した処理を行う。

**【0065】**

まず、訳語生成部600は、対象のフレーズを翻訳して、フレーズ訳の候補となるフレーズ訳候補を1以上生成する(S800)。次に、ページ検索部610は、フレーズ訳候補のそれぞれについて、当該フレーズ訳候補を含むネットワーク上のページを検索する(S810)。そして、フレーズ訳生成部620は、それぞれのフレーズ訳候補を含むページがヒットしたか否かに基づいて、当該フレーズ訳候補をフレーズ訳とするか否かを選択する(S820)。

**【0066】**

例えば、対象のフレーズが「enterprise software」の場合、訳語生成部600は、フレーズ訳候補として、「計画ソフトウェア」、「事業ソフトウェア」、及び「企業ソフトウェア」を生成する。次に、ページ検索部610は、フレーズ訳候補である「計画ソフトウェア」を含むページ、「事業ソフトウェア」を含むページ、及び「企業ソフトウェア」を含むページを検索する。そして、フレーズ訳生成部620は、「計画ソフトウェア」を含むページ及び「事業ソフトウェア」を含むページがヒットせず、「企業ソフトウェア」を含むページがヒットした場合に、「企業ソフトウェア」をフレーズ訳とする。

**【0067】**

ここで、フレーズ訳生成部620は、複数のフレーズ訳候補について、各フレ

ーズ訳候補を含むページがヒットした場合、最も多くのページがヒットしたフレーズ訳候補をフレーズ訳として選択してよい。これに代えて、フレーズ訳生成部 620 は、ネットワーク上のページにヒットした回数が最も多かったフレーズ訳候補を、フレーズ訳として選択してよい。

#### 【0068】

また、フレーズ翻訳部 180 は、上記の処理を以下の方法により行ってもよい。

まず、訳語生成部 600 は、分野別辞書 175 a～b を参照して、対象のフレーズに含まれるそれぞれの語を翻訳し、それぞれの語に対応する訳語を 1 以上生成する (S800)。次に、ページ検索部 610 は、対象のフレーズに含まれる全ての語について、当該語に対応する訳語を少なくとも 1 つずつを含むネットワーク上のページを検索する (S810)。次に、フレーズ訳生成部 620 は、検索したページにおいて、フレーズに含まれる全ての語について、当該語に対応する訳語の少なくとも 1 つずつを含む箇所の語句に基づいて、フレーズ訳を生成する (S820)。

#### 【0069】

例えば、対象のフレーズが「enterprise software」の場合、訳語生成部 600 は、フレーズに含まれる「enterprise」及び「software」の語を翻訳し、「enterprise」に対応する訳語 {計画、事業、企業} と、「software」に対応する訳語 {ソフトウェア} とを生成する。次に、ページ検索部 610 は、対象フレーズに含まれるそれぞれの語について、当該語に対応する訳語を 1 つずつ含むネットワーク上のページ、すなわち「計画」と「ソフトウェア」とを含むページ A、「事業」と「ソフトウェア」とを含むページ B、「企業」と「ソフトウェア」とを含むページ C をそれぞれ検索する。そして、フレーズ訳生成部 620 は、ページ A から C において、フレーズに含まれる全ての語の訳語を少なくとも 1 つずつ含む箇所の語句に基づき、フレーズ訳を生成する。例えばこの場合において、ページ A 及びページ B の中で、「計画ソフトウェア」及び「事業ソフトウェア」と記載された箇所がなく、ページ C の中で「企業ソフトウェア」と記載された箇所がある場合、ページ C

において全ての語の訳語を少なくとも1つずつ含む箇所「企業ソフトウェア」を、「enterprise software」のフレーズ訳とする。

#### 【0070】

以上において、フレーズ訳生成部620は、検索した結果ヒットしたページの数に基づいて、フレーズ訳を生成してもよい。すなわち例えば、上記の例において、「計画」と「ソフトウェア」とを含むページのヒット数、「事業」と「ソフトウェア」とを含むページのヒット数、及び「企業」と「ソフトウェア」を含むページのヒット数のうち、最もヒット数が多いページに対応する語を選択し、フレーズ訳を生成してもよい。

#### 【0071】

図10は、本実施形態に係るドキュメント翻訳部120及びフレーズ翻訳部180において、登録対象のフレーズが名詞句「Visitor reviews」であった場合における翻訳結果の一例を示す。

#### 【0072】

図10(a)は、ドキュメント翻訳部120がドキュメントにおける名詞句以外の部分を翻訳する場合において、文を優先して翻訳した場合における翻訳結果を示す。

#### 【0073】

まず、形態素解析部210は、対象の名詞句の形態素解析を行い、それぞれの語の品詞等を解析する。次に、構文解析部230は、分野別辞書175a～bに登録された文法に基づいて、構文解析を行なう。

#### 【0074】

構文解析において、構文解析部230は、各英単語に対応付けて、当該英単語の品詞毎に、当該品詞として使用される頻度の低さを示すコストを割り当てる。例えば、英単語「Visitor」が名詞として用いられるコストは、()内に図示した通り5である。

#### 【0075】

次に、構文解析部230は、分野別辞書175a～bに登録された文法に記述された組合せを用いて句を生成し、当該句のコストを割り当てる。本例において

、名詞句+名詞句として用いられる場合のコストは80であり、名詞単体で名詞句として用いられる場合のコストは18であり、動詞単体で動詞句として用いられるコストは15である。

#### 【0076】

次に、構文解析部230は、句を組み合わせて完成文を生成し、完成文のコストを割り当てる。本例において、名詞句+動詞句で文を構成するコストは18であり、名詞句単体の完成文990a、及び名詞句+動詞句による完成文990bのコストは共に200である。

#### 【0077】

次に、構文解析部230は、以上により解析されたそれぞれの完成文990a～bについて、コストの合計値を算出する。例えば、完成文990aのコストの合計値は、「名詞(5)+名詞(5)+名詞句(80)+完成文(200)=290」である。一方、完成文990bのコストの合計値は261である。

#### 【0078】

以上の構文解析の結果、構文解析部230は、コスト合計値が最も小さい文法、すなわち「Visitor reviews」を完成文990bに翻訳する文法を、「Visitor reviews」の構文解析結果として出力する。この結果、ドキュメント訳生成部240は、翻訳結果「訪問者はレビューする」を出力する。

#### 【0079】

図10(b)は、フレーズ翻訳部180が名詞句を優先して翻訳する場合における翻訳結果を示す。フレーズ翻訳部180は、名詞句訳を生成する場合において、ドキュメント翻訳部120がドキュメントにおける名詞句以外の部分を翻訳する場合と比較し、名詞句として翻訳する文法規則を使用する優先度をより高く設定する。すなわち、図10(b)に示した様に、図10(a)における名詞句のみからなる完成文のコストを、完成文990bにおける完成文のコストより所定の値、例えば150減じる。これにより、構文解析部616は、「Visitor reviews」を完成文990aに翻訳する文法を、「Visitor reviews」の構文解析結果として出力する。この結果、フレーズ訳生成

部 620 は、翻訳結果「訪問者レビュー」を出力する。

【0080】

以上に示した様に、フレーズ翻訳部 180 は、名詞句訳を生成する場合において、名詞句以外の部分を翻訳する場合と比較し、名詞句として優先して翻訳する文法規則を優先する。より具体的には、フレーズ翻訳部 180 は、登録すべき名詞句を翻訳する場合において、名詞句として優先して翻訳する文法規則を、名詞と動詞の組からなる文に翻訳する文法規則に対して相対的に優先度を高める。

【0081】

フレーズ翻訳部 180 は、名詞句として優先的に翻訳する方法として、フレーズ翻訳部 180 が名詞句を翻訳する場合に用いる名詞句用の文法規則を分野別辞書 175 a～b の少なくとも 1 つに登録しておいてもよい。

【0082】

以上に示したフレーズ翻訳部 180 によれば、対象のドキュメントから抽出された名詞句の名詞句訳を生成する場合において、ドキュメントにおける名詞句以外の部分を翻訳する場合と比較し、名詞句として翻訳する文法規則を使用する優先度をより高く設定する。これにより、フレーズ翻訳部 180 は、抽出した名詞句に適した翻訳を行なうことができ、翻訳の精度を高めることができる。

【0083】

図 11 は、本実施形態に係るコンピュータ 1000 のハードウェア構成の一例を示す。本実施形態に係る翻訳フロントエンドシステム 100 及び／又は辞書更新サーバ 160 は、ホスト・コントローラ 1182 により相互に接続される CPU 1100、RAM 1120、グラフィック・コントローラ 1175、及び表示装置 1180 を有する CPU 周辺部と、入出力コントローラ 1184 によりホスト・コントローラ 1182 に接続される通信インターフェイス 1130、ハードディスク・ドライブ 1140、及び CD-ROM ドライブ 1160 を有する入出力部と、入出力コントローラ 1184 に接続される ROM 1110、フレキシブルディスク・ドライブ 1150、及び入出力チップ 1170 を有するレガシー入出力部とを備えるコンピュータ 1000 により実現される。

【0084】

ホスト・コントローラ 1182 は、RAM 1120 と、高い転送レートで RAM 1120 をアクセスする CPU 1100 及びグラフィック・コントローラ 1175 とを接続する。CPU 1100 は、ROM 1110 及び RAM 1120 に格納されたプログラムに基づいて動作し、各部の制御を行う。グラフィック・コントローラ 1175 は、CPU 1100 等が RAM 1120 内に設けたフレーム・バッファ上に生成する画像データを取得し、表示装置 1180 上に表示させる。これに代えて、グラフィック・コントローラ 1175 は、CPU 1100 等が生成する画像データを格納するフレーム・バッファを、内部に含んでもよい。

#### 【0085】

入出力コントローラ 1184 は、ホスト・コントローラ 1182 と、比較的高速な入出力装置である通信インターフェイス 1130、ハードディスク・ドライブ 1140、CD-ROM ドライブ 1160 を接続する。通信インターフェイス 1130 は、ネットワークを介して他の装置と通信する。ハードディスク・ドライブ 1140 は、コンピュータ 1000 が使用するプログラム及びデータを格納する。CD-ROM ドライブ 1160 は、CD-ROM 1195 からプログラム又はデータを読み取り、RAM 1120 及び／又はハードディスク・ドライブ 1140 に提供する。

#### 【0086】

また、入出力コントローラ 1184 には、ROM 1110 と、フレキシブルディスク・ドライブ 1150 や入出力チップ 1170 等の比較的低速な入出力装置とが接続される。ROM 1110 は、コンピュータ 1000 が起動時に実行するブート・プログラムや、コンピュータ 1000 のハードウェアに依存するプログラム等を格納する。フレキシブルディスク・ドライブ 1150 は、フレキシブルディスク 1190 からプログラム又はデータを読み取り、入出力コントローラ 1184 を介して CPU 1100 及び／又はハードディスク・ドライブ 1140 に提供する。入出力チップ 1170 は、フレキシブルディスク 1190 や、例えばパラレル・ポート、シリアル・ポート、キーボード・ポート、マウス・ポート等を介して各種の入出力装置を接続する。

#### 【0087】

RAM1120を介してCPU1100に提供されるプログラムは、フレキシブルディスク1190、CD-ROM1195、又はICカード等の記録媒体に格納されて利用者によって提供される。プログラムは、記録媒体から読み出され、入出力コントローラ1184及びRAM1120を介してコンピュータ1000にインストールされ、CPU1100において実行される。

#### 【0088】

コンピュータ1000にインストールされて実行され、コンピュータ1000を翻訳フロントエンドシステム100として機能させるプログラムは、辞書管理モジュールと、形態素解析モジュールと、フレーズ抽出モジュールと、構文解析モジュールと、ドキュメント訳生成モジュールと、ドキュメント分野選択モジュールとを含むドキュメント翻訳モジュールを備える。これらのプログラム又はモジュールは、コンピュータ1000を、辞書管理部200と、形態素解析部210と、フレーズ抽出部220と、構文解析部230と、ドキュメント訳生成部240と、ドキュメント分野選択部250とを含むドキュメント翻訳部120としてそれぞれ機能させる。また、翻訳辞書記録部110及び抽出フレーズ記録部125は、ハードディスク・ドライブ1140又はネットワークに接続されたサーバ上の記録媒体として実現されてもよい。

#### 【0089】

コンピュータ1000にインストールされて実行され、コンピュータ1000を辞書更新サーバ160として機能させるプログラムは、登録フレーズ選択モジュールと、分野別出現頻度算出モジュール及び登録先分野選択モジュールとを含む登録先選択モジュール、訳語生成モジュールと、ページ検索モジュールと、形態素解析モジュールと、構文解析モジュールと、フレーズ訳生成モジュールとを含むフレーズ翻訳モジュール、及び、辞書登録モジュールを備える。これらのプログラム又はモジュールは、コンピュータ1000を、登録フレーズ選択部400と、分野別出現頻度算出部420及び登録先分野選択部430とを含む登録先選択部410、訳語生成部600と、ページ検索部610と、形態素解析部613と、構文解析部616と、フレーズ訳生成部620とを含むフレーズ翻訳部180、及び、辞書登録部190としてそれぞれ機能させる。また、登録フレーズ

記録部 140、翻訳辞書記録部 170 及び更新辞書 185 は、ハードディスク・ドライブ 1140 又はネットワークに接続されたサーバ上の記録媒体として実現されてもよい。

#### 【0090】

以上に示したプログラム又はモジュールは、外部の記憶媒体に格納されてもよい。記憶媒体としては、フレキシブルディスク 1190、CD-ROM 1195 の他に、DVD や PD 等の光学記録媒体、MD 等の光磁気記録媒体、テープ媒体、IC カード等の半導体メモリ等を用いることができる。また、専用通信ネットワークやインターネットに接続されたサーバシステムに設けたハードディスク又は RAM 等の記憶装置を記録媒体として使用し、ネットワークを介してプログラムをコンピュータ 1000 に提供してもよい。

#### 【0091】

以上、本発明を実施形態を用いて説明したが、本発明の技術的範囲は上記実施形態に記載の範囲には限定されない。上記実施形態に、多様な変更または改良を加えることができる。そのような変更または改良を加えた形態も本発明の技術的範囲に含まれ得ることが、特許請求の範囲の記載から明らかである。

#### 【0092】

以上に説明した実施形態によれば、以下の各項目に示す翻訳システム、辞書更新サーバ、翻訳方法、及び、これらのプログラムと記録媒体が実現される。

#### 【0093】

(項目 1) ドキュメントを翻訳する翻訳システムであって、予め定められた分野別に分類された複数の分野別辞書を管理する辞書管理部と、前記ドキュメントから、名詞句 (Noun Phrase) を抽出するフレーズ抽出部と、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部とを備える翻訳システム。

(項目 2) 前記ドキュメントを翻訳したときに前記複数の分野別辞書のそれぞ



れを用いた頻度に基づいて、前記ドキュメントの分野を選択するドキュメント分野選択部を更に備え、前記登録先選択部は、前記ドキュメント分野選択部により選択された分野に基づいて、抽出された前記名詞句を登録すべき分野を選択する項目 1 記載の翻訳システム。

#### 【 0 0 9 4 】

(項目 3) 前記ドキュメント分野選択部は、複数の前記ドキュメントのそれぞれの翻訳において、前記複数の分野別辞書のそれぞれを用いた頻度に基づいて、前記複数のドキュメントのそれぞれの分野を選択し、前記フレーズ抽出部は、前記複数のドキュメントから、前記名詞句を抽出し、前記登録先選択部は、前記複数のドキュメントのそれぞれについて、前記名詞句が出現した頻度及び当該ドキュメントの分野に基づいて、前記名詞句を登録すべき分野を選択する項目 2 記載の翻訳システム。

(項目 4) 前記複数のドキュメントにおいて前記名詞句が出現した頻度が予め定められた下限値未満の場合に、前記名詞句及び前記名詞句訳の組を、前記複数の分野別辞書のいずれにも登録させないことを選択する登録フレーズ選択部を更に備える項目 3 記載の翻訳システム。

#### 【 0 0 9 5 】

(項目 5) 一の前記分野別辞書は、他の複数の前記分野別辞書に対応する複数の分野のいずれかに分類されない語句を登録する基本辞書であり、前記登録先選択部は、前記複数のドキュメントのそれぞれについて前記名詞句が出現した頻度及び当該ドキュメントの分野に基づいて、前記複数の分野のそれぞれに対する前記名詞句の出現頻度を算出する分野別出現頻度算出部と、前記複数の分野のそれぞれに対する前記出現頻度に基づいて、前記名詞句及び前記名詞句訳の組を、前記複数の分野別辞書及び前記基本辞書のいずれに登録するかを選択する登録先分野選択部とを有し、前記辞書登録部は、前記登録先分野選択部により前記名詞句を前記基本辞書に登録することを選択された場合に、前記名詞句及び前記名詞句訳の組を前記基本辞書に登録する項目 2 記載の翻訳システム。

#### 【 0 0 9 6 】

(項目 6) 前記登録先選択部は、前記複数のドキュメントに対応する複数の分

野のそれぞれに対して前記名詞句が出現した度合に基づいて、前記名詞句を登録すべき分野を選択する項目 2 記載の翻訳システム。

(項目 7) 前記翻訳部は、前記登録先選択部により選択された、前記名詞句を登録すべき分野に対応する前記分野別辞書を優先して前記名詞句を翻訳し、前記名詞句訳を生成する項目 1 記載の翻訳システム。

(項目 8) 前記翻訳部は、前記名詞句訳を生成する場合において、前記ドキュメントにおける名詞句以外の部分を翻訳する場合と比較し、名詞句として翻訳する文法規則を使用する優先度をより高く設定する項目 1 記載の翻訳システム。

#### 【0097】

(項目 9) 前記翻訳部は、前記名詞句を翻訳して、前記名詞句訳の候補となる名詞句訳候補を生成する訳語生成部と、前記名詞句訳候補を含むネットワーク上のページを検索するページ検索部と、前記名詞句訳候補を含むページがヒットしたか否かに基づいて、前記名詞句訳候補を前記名詞句訳とするか否かを選択する名詞句訳生成部とを有する項目 1 記載の翻訳システム。

#### 【0098】

(項目 10) 予め定められた分野別に分類された複数の分野別辞書を管理する端末においてドキュメントの翻訳に用いる辞書を更新する辞書更新サーバであって、前記ドキュメントから抽出された名詞句を前記端末から受信する名詞句受信部と、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、前記端末から受信した前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、前記前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部とを備える辞書更新サーバ。

#### 【0099】

(項目 11) ドキュメントを翻訳し、翻訳に用いる辞書を外部の辞書更新サーバの指示に基づき更新する端末を有する翻訳システムであって、前記端末は、予め定められた分野別に分類された複数の分野別辞書を格納する翻訳辞書記録部と、前記ドキュメントから名詞句を抽出するフレーズ抽出部と、前記複数の分野別辞書を用いて前記ドキュメントを翻訳するドキュメント翻訳部とを備え、前記辞

書更新サーバは、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句および前記名詞句訳の組の登録を指示する辞書登録部とを備え、前記翻訳辞書記録部は、前記辞書登録部の登録指示に基づき、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する翻訳システム。

#### 【0100】

(項目12) ドキュメントを翻訳する翻訳システム用のプログラムであって、前記翻訳システムを、予め定められた分野別に分類された複数の分野別辞書を管理する辞書管理部と、前記ドキュメントから、名詞句(Noun Phrase)を抽出するフレーズ抽出部と、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部として機能させるプログラム。

#### 【0101】

(項目13) 予め定められた分野別に分類された複数の分野別辞書を管理する端末においてドキュメントの翻訳に用いる辞書を更新する辞書更新サーバ用のプログラムであって当該プログラムは、当該辞書更新サーバを、前記ドキュメントから抽出された名詞句を前記端末から受信する名詞句受信部と、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を選択する登録先選択部と、前記端末から受信した前記名詞句を翻訳して前記名詞句の訳である名詞句訳を生成する翻訳部と、前記登録先選択部により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を登録する辞書登録部として機能させるプログラム。

#### 【0102】

(項目 14) 項目 12 から項目 13 のいずれかに記載のプログラムを記録した記録媒体。

(項目 15) コンピュータによりドキュメントを翻訳する翻訳システムにおける翻訳方法であって、予め定められた分野別に分類された複数の分野別辞書を、コンピュータにより管理させる辞書管理段階と、前記ドキュメントから、名詞句 (Noun Phrase) を、コンピュータにより抽出させるフレーズ抽出段階と、前記複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された前記名詞句を登録すべき分野を、コンピュータにより選択させる登録先選択段階と、前記名詞句を翻訳して前記名詞句の訳である名詞句訳を、コンピュータにより生成させる翻訳段階と、前記登録先選択段階により選択された分野に対応する前記分野別辞書に、前記名詞句及び前記名詞句訳の組を、コンピュータにより登録させる辞書登録段階とを備える翻訳方法。

#### 【0103】

##### 【発明の効果】

上記説明から明らかなように、本発明によれば、翻訳対象となるドキュメントから抽出したフレーズの訳を生成し、翻訳辞書に登録することにより、新たに作り出された語句等の増加によって翻訳精度が低下するのを防ぐことができる。

##### 【図面の簡単な説明】

【図 1】 本発明の実施形態に係る翻訳システム 10 の構成を示す。

【図 2】 本発明の実施形態に係る翻訳辞書記録部 110 及び翻訳辞書記録部 170 に格納される翻訳辞書 117 及び翻訳辞書 177 の階層構造の一例を示す。

【図 3】 本発明の実施形態に係るドキュメント翻訳部 120 の構成を示す。

【図 4】 本発明の実施形態に係るドキュメント翻訳部 120 の処理フローを示す。

【図 5】 本発明の実施形態に係るフレーズ分類部 130 の構成を示す。

【図 6】 本発明の実施形態に係るフレーズ分類部 130 の処理フローを示す。

【図 7】 本発明の実施形態に係るフレーズ翻訳部 180 の構成を示す。

【図 8】 本発明の実施形態に係るフレーズ翻訳部 180 の処理フローを示す。

【図 9】 本発明の実施形態に係るフレーズ翻訳部 180 の、ネットワークを介したフレーズ訳生成処理フローを示す。

【図 10】 本発明の実施形態に係るドキュメント翻訳部 120 及びフレーズ翻訳部 180 における翻訳結果の一例を示す。図 10 (a) は、文を優先して翻訳する場合における翻訳結果を示す。図 10 (b) は、名詞句を優先して翻訳する場合における翻訳結果を示す。

【図 11】 本発明の実施形態に係るコンピュータ 1000 のハードウェア構成の一例を示す。

【符号の説明】

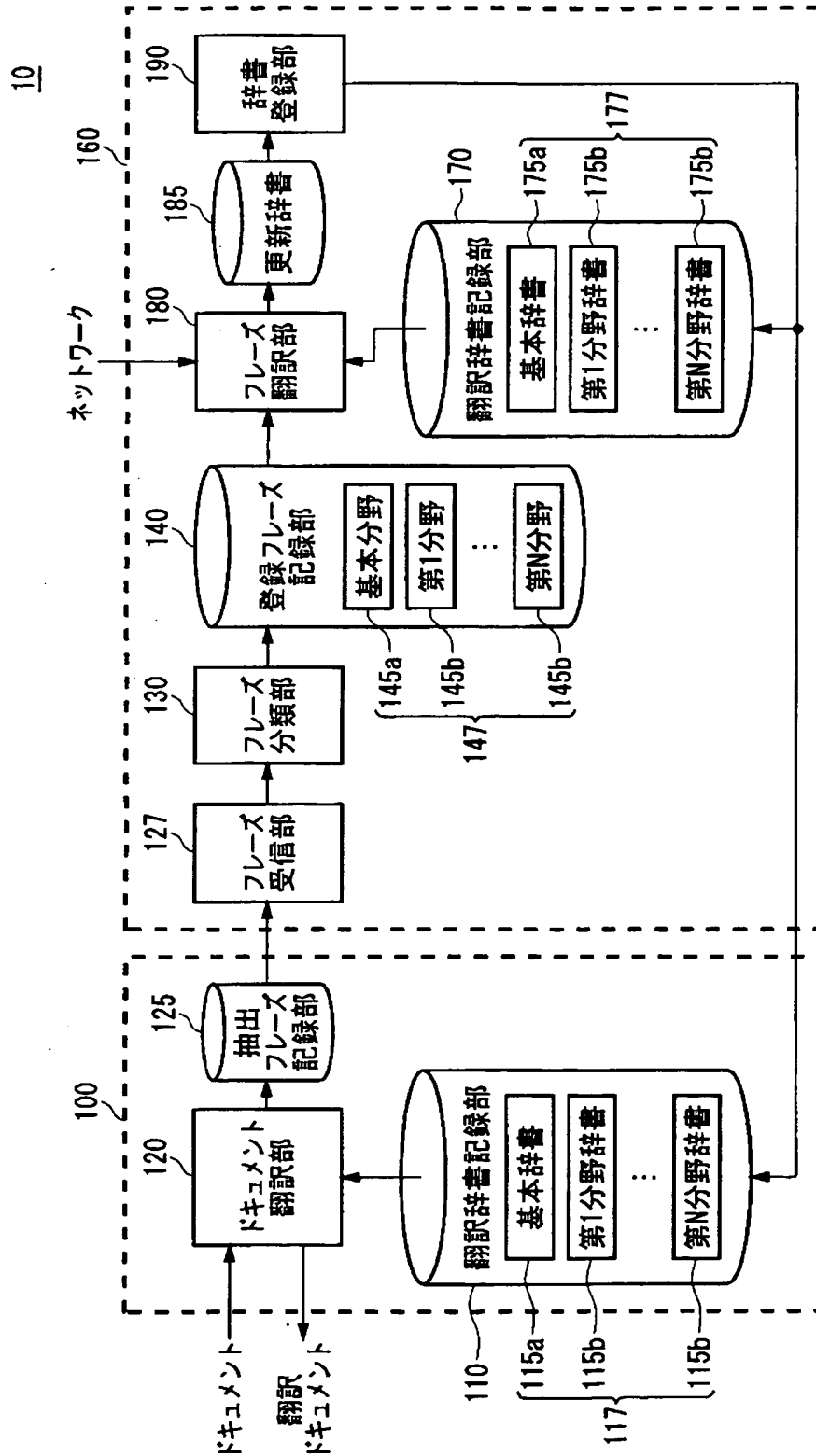
- 10 翻訳システム
- 100 翻訳フロントエンドシステム
- 110 翻訳辞書記録部
- 115 a～b 分野別辞書
- 117 翻訳辞書
- 120 ドキュメント翻訳部
- 125 抽出フレーズ記録部
- 127 フレーズ受信部
- 130 フレーズ分類部
- 140 登録フレーズ記録部
- 145 a～b 分野別登録フレーズ記録ファイル
- 160 辞書更新サーバ
- 170 翻訳辞書記録部
- 175 a～b 分野別辞書
- 177 翻訳辞書
- 180 フレーズ翻訳部
- 185 更新辞書

1 9 0 辞書登録部  
2 0 0 辞書管理部  
2 1 0 形態素解析部  
2 2 0 フレーズ抽出部  
2 3 0 構文解析部  
2 4 0 ドキュメント訳生成部  
2 5 0 ドキュメント分野選択部  
4 0 0 登録フレーズ選択部  
4 1 0 登録先選択部  
4 2 0 分野別出現頻度算出部  
4 3 0 登録先分野選択部  
6 0 0 訳語生成部  
6 0 5 優先度設定部  
6 1 0 ページ検索部  
6 1 3 形態素解析部  
6 1 6 構文解析部  
6 2 0 フレーズ訳生成部  
9 0 0 辞書  
9 1 0 分野別辞書  
9 0 5 基本辞書  
9 1 5 主分野  
9 2 0 主分野別辞書  
9 2 5 副分野別辞書  
9 9 0 a ~ b 完成文  
1 0 0 0 コンピュータ  
1 1 0 0 C P U  
1 1 1 0 R O M  
1 1 2 0 R A M  
1 1 3 0 通信インターフェイス

- 1140 ハードディスク・ドライブ
- 1150 フレキシブルディスク・ドライブ
- 1160 CD-ROMドライブ
- 1170 入出力チップ
- 1175 グラフィック・コントローラ
- 1180 表示装置
- 1182 ホスト・コントローラ
- 1184 入出力コントローラ
- 1190 フレキシブルディスク
- 1195 CD-ROM

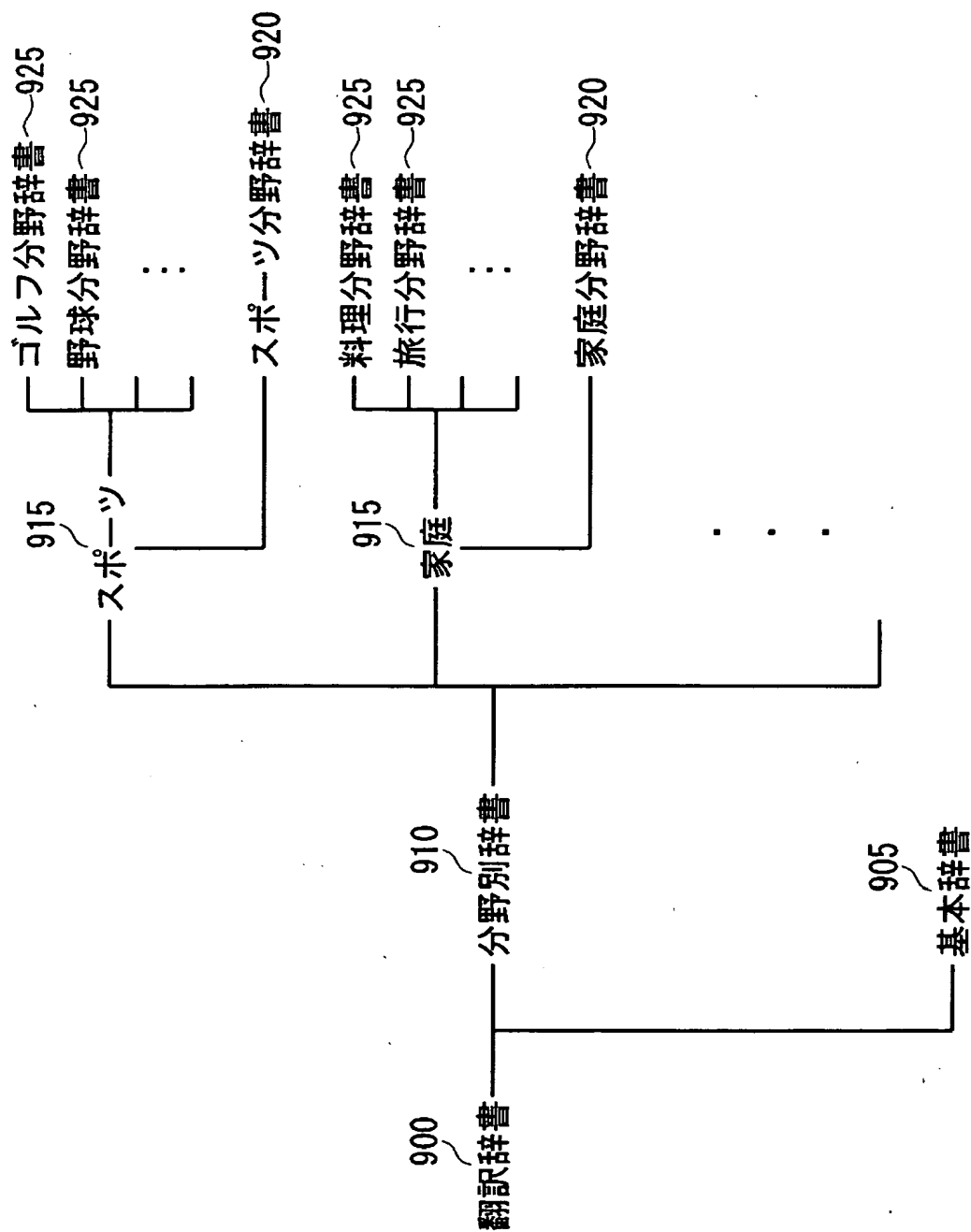
【書類名】 図面

【図 1】

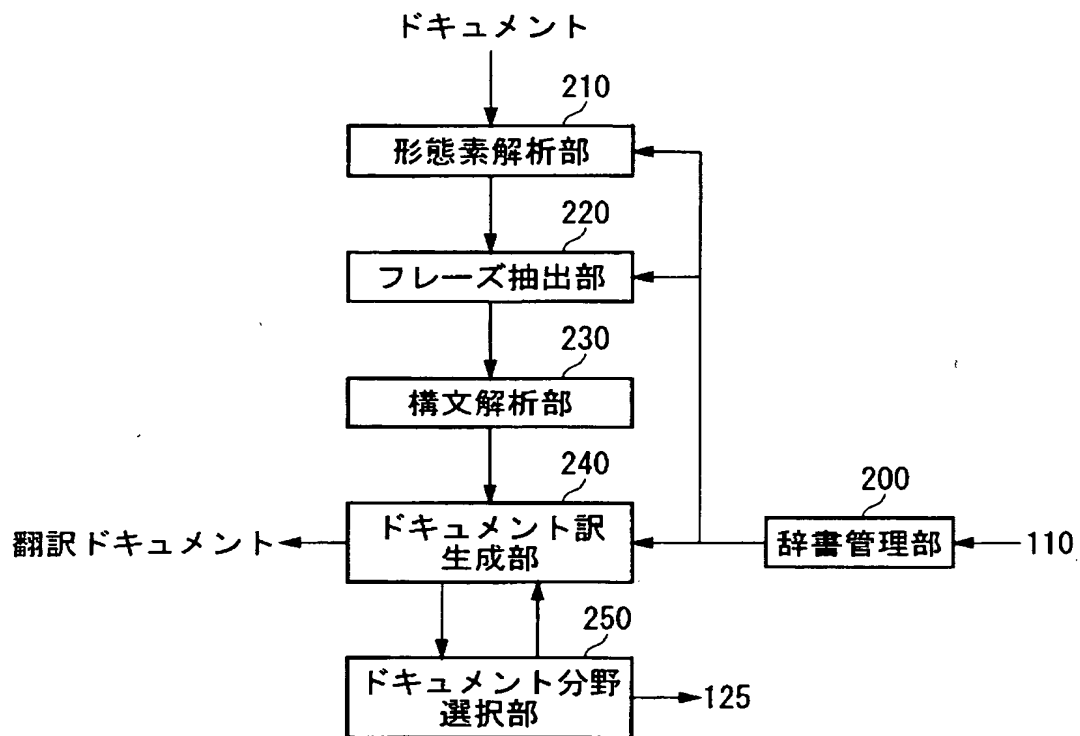




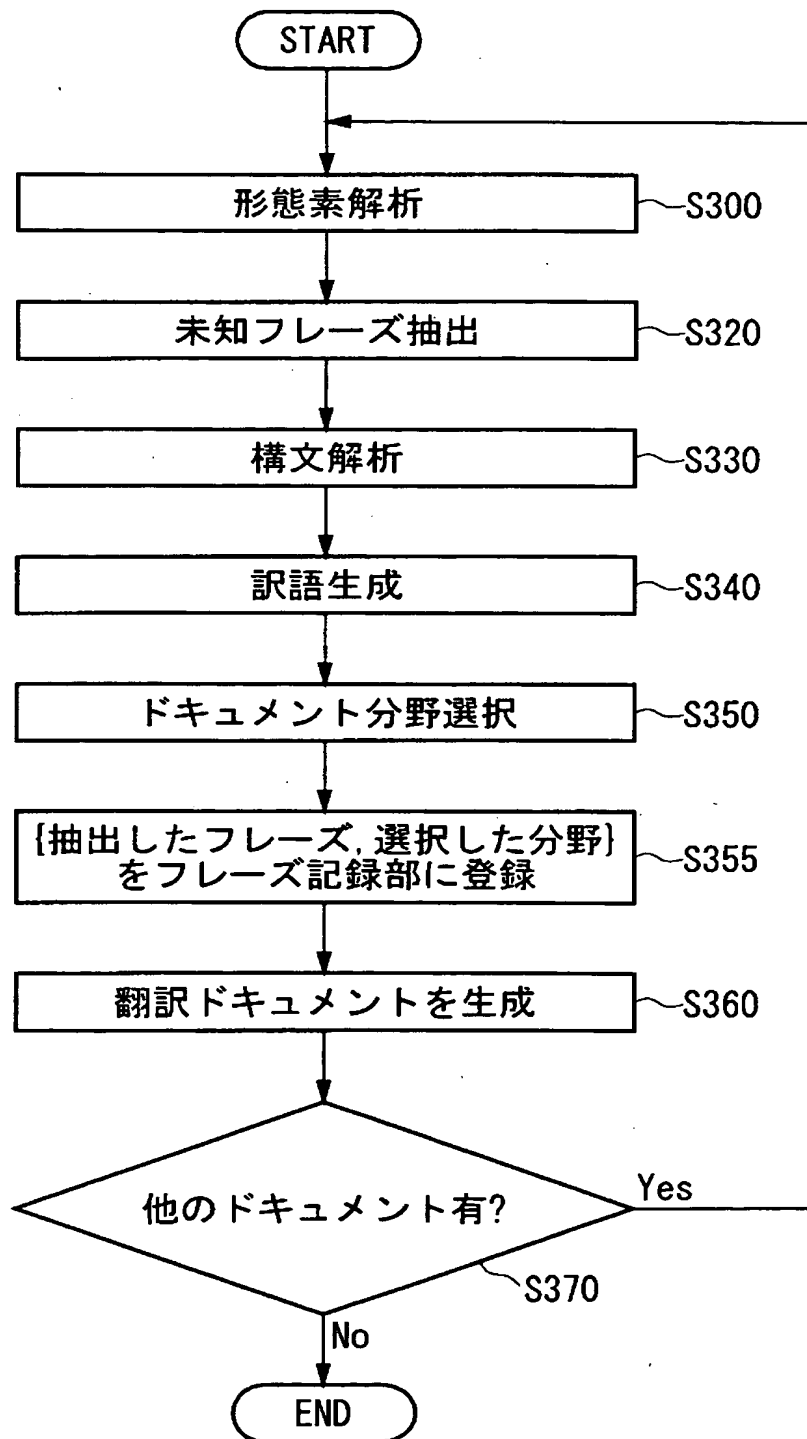
【図 2】



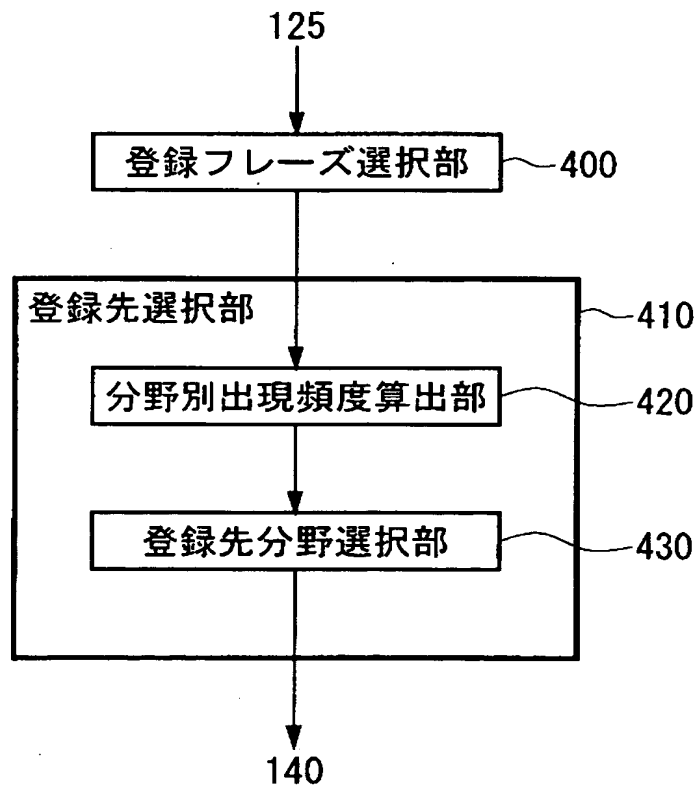
【図 3】

120

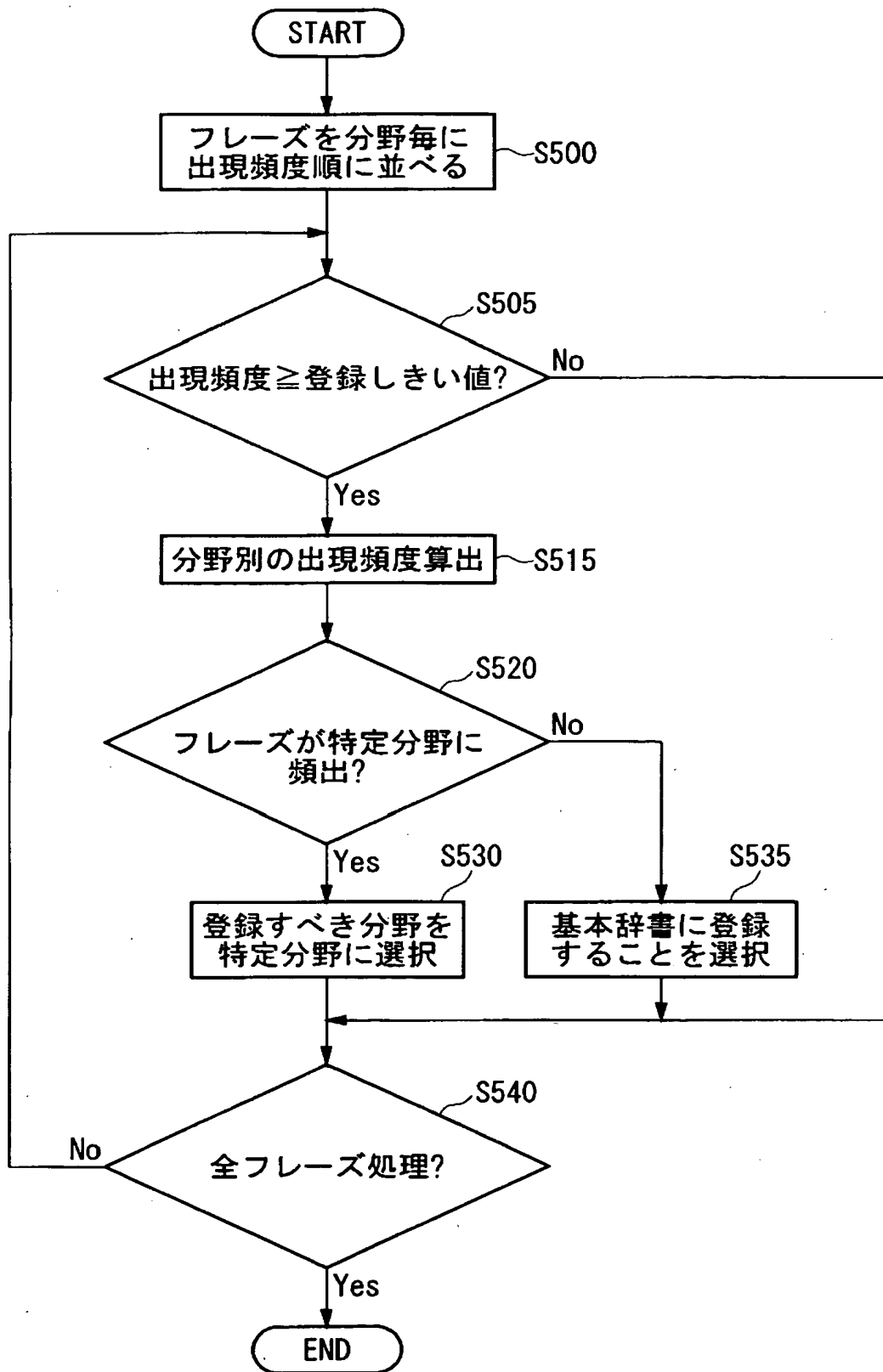
【図 4】



【図 5】

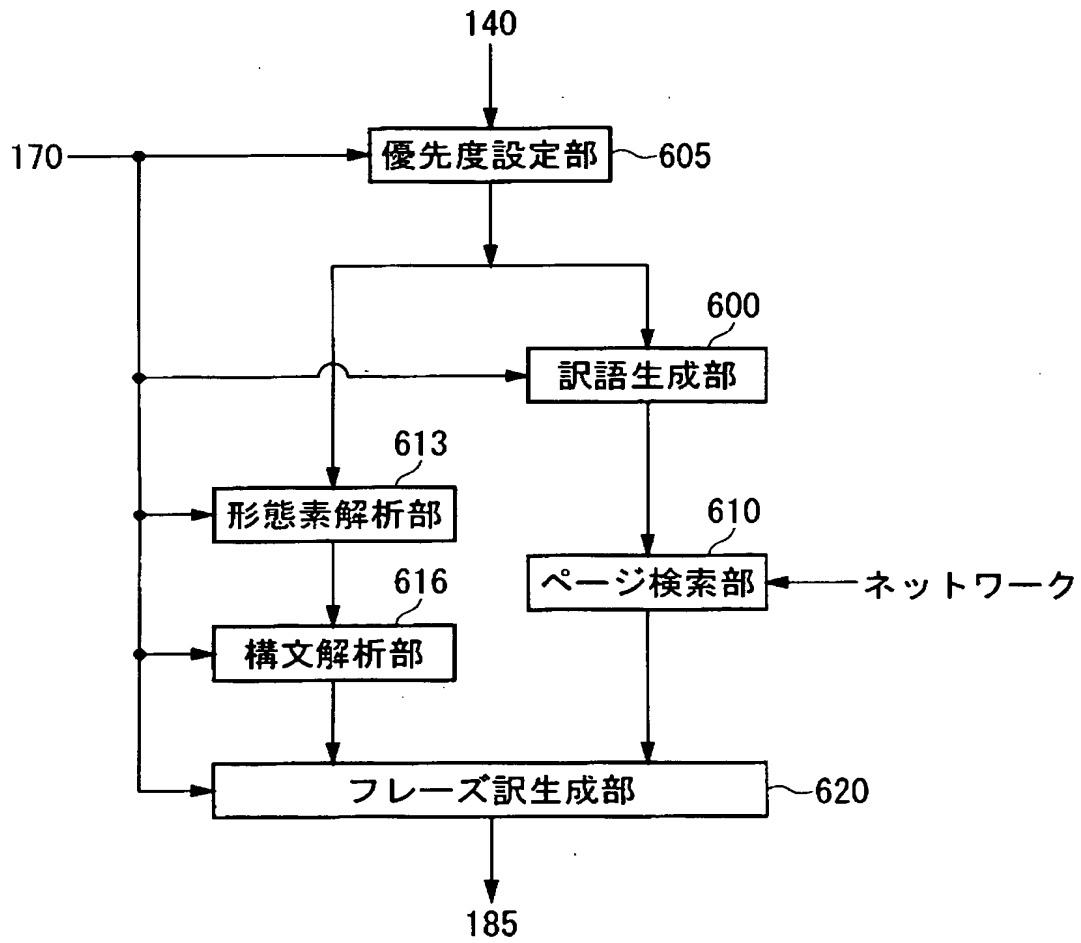
130

【図 6】

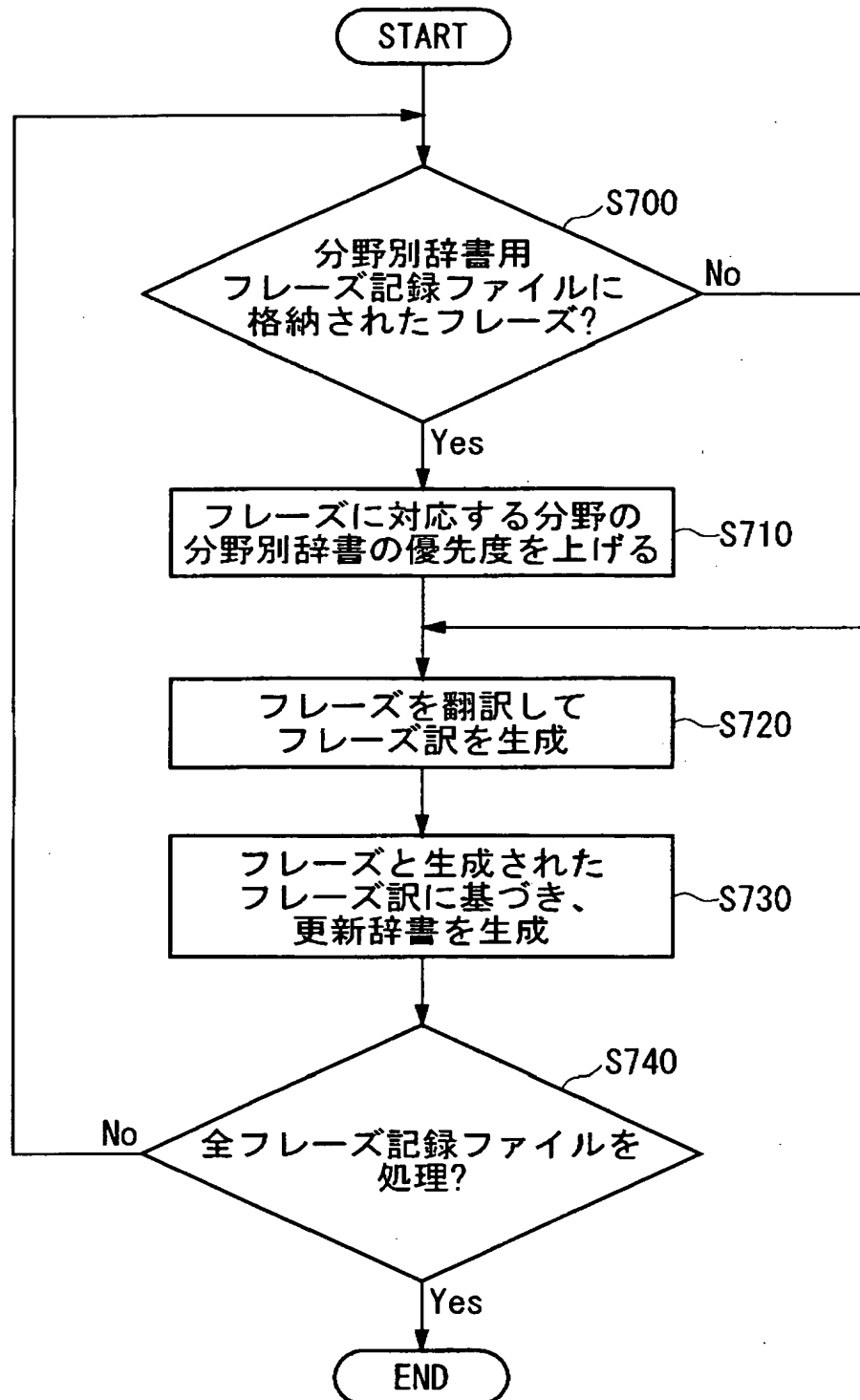


【図 7】

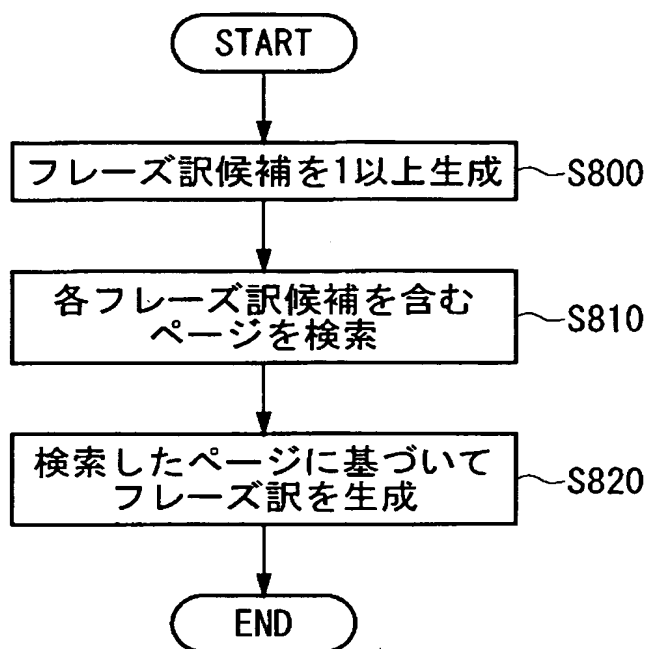
180



【図 8】



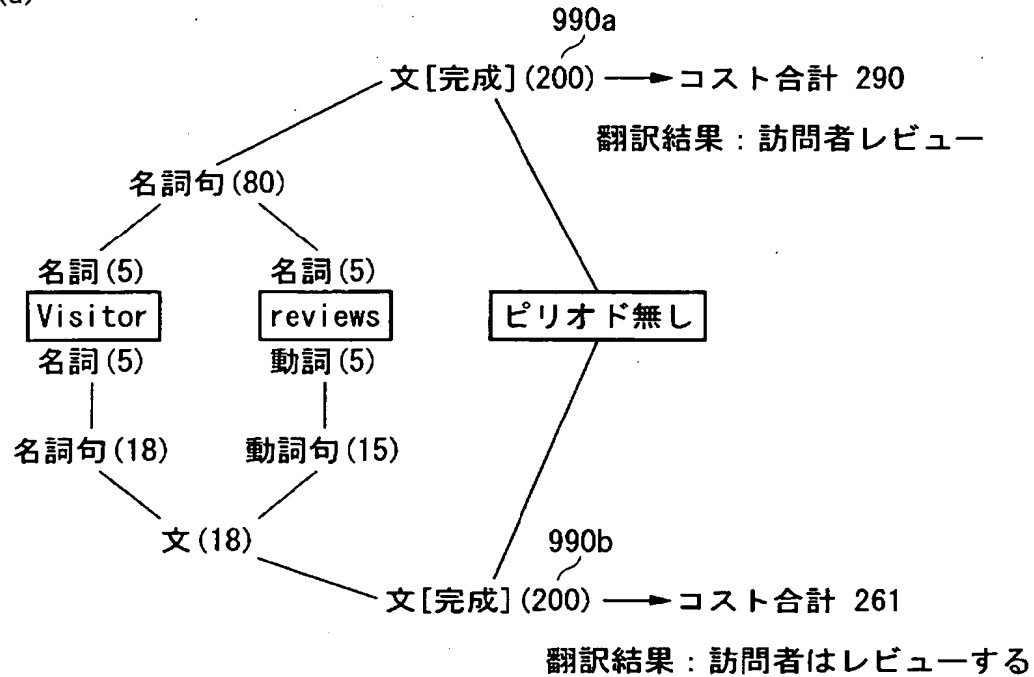
【図 9】

S720

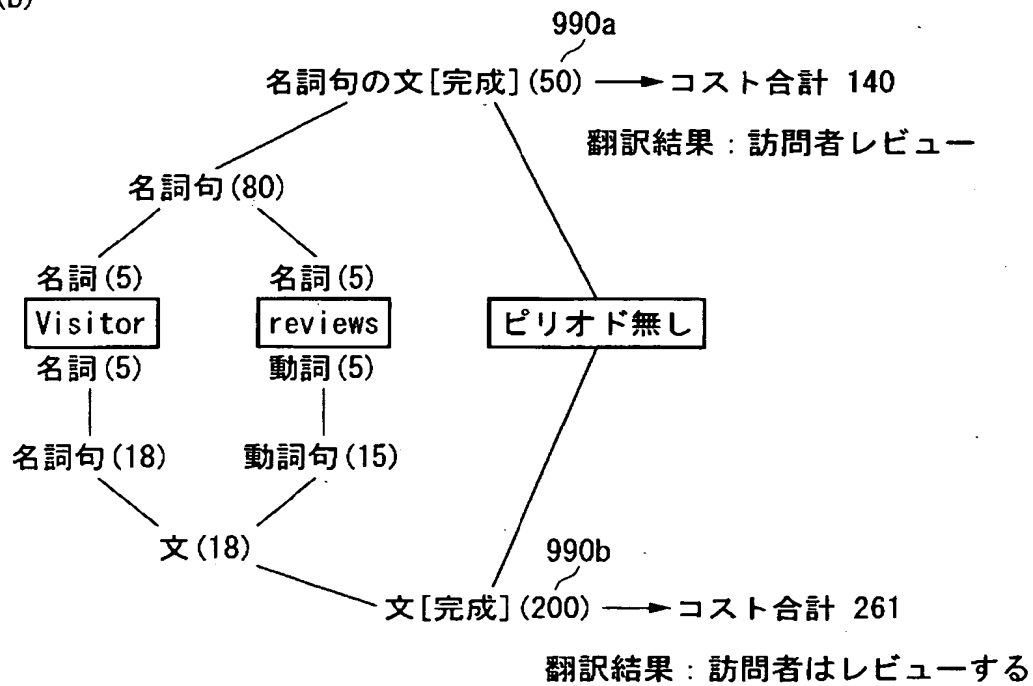


【図 10】

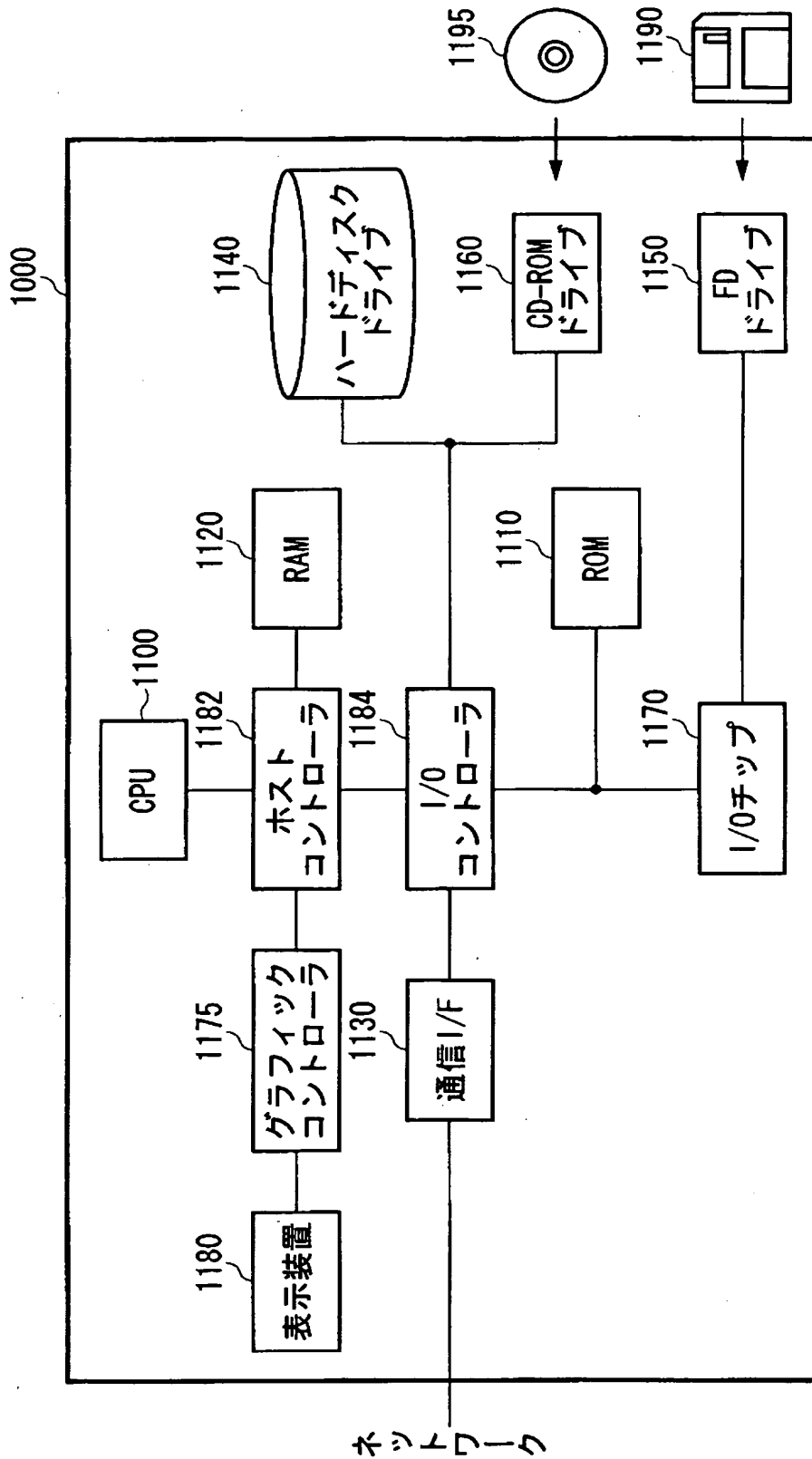
(a)



(b)



【図 11】



【書類名】 要約書

【要約】

【課題】 新たに作り出された語句等の増加によって翻訳精度が低下するのを防ぐことができる翻訳システムを提供する。

【解決手段】 ドキュメントを翻訳する翻訳システムは、予め定められた分野別に分類された複数の分野別辞書を管理する辞書管理部と、ドキュメントから、名詞句を抽出するフレーズ抽出部と、複数の分野別辞書にそれぞれ対応する複数の分野のうち、抽出された名詞句を登録すべき分野を選択する登録先選択部と、名詞句を翻訳して名詞句の訳である名詞句訳を生成する翻訳部と、登録先選択部により選択された分野に対応する分野別辞書に、名詞句及び名詞句訳の組を登録する辞書登録部とを備える。

【選択図】 図 1

## 認定・付加情報

特許出願の番号 特願 2003-072324  
受付番号 50300433929  
書類名 特許願  
担当官 末武 実 1912  
作成日 平成 15 年 4 月 25 日

## &lt;認定情報・付加情報&gt;

## 【特許出願人】

【識別番号】 390009531  
【住所又は居所】 アメリカ合衆国 10504、ニューヨーク州 アーモンク ニュー オーチャード ロード  
【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

## 【代理人】

【識別番号】 100086243  
【住所又は居所】 神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内  
【氏名又は名称】 坂口 博

## 【代理人】

【識別番号】 100091568  
【住所又は居所】 神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内  
【氏名又は名称】 市位 嘉宏

## 【代理人】

【識別番号】 100108501  
【住所又は居所】 神奈川県大和市下鶴間 1623 番 14 日本アイ・ビー・エム株式会社 知的所有権  
【氏名又は名称】 上野 剛史

## 【復代理人】

申請人  
【識別番号】 100104156  
【住所又は居所】 東京都新宿区新宿 1 丁目 24 番 12 号 東信ビル 6 階 龍華国際特許事務所  
【氏名又は名称】 龍華 明裕

次頁無

特願 2003-072324

出願人履歴情報

識別番号

[390009531]

1. 変更年月日

2000年 5月16日

[変更理由]

名称変更

住 所

アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)

氏 名

インターナショナル・ビジネス・マシーンズ・コーポレーション

2. 変更年月日

2002年 6月 3日

[変更理由]

住所変更

住 所

アメリカ合衆国10504、ニューヨーク州 アーモンク ニュー オーチャード ロード

氏 名

インターナショナル・ビジネス・マシーンズ・コーポレーション